Bayes factors for linear regression

Rui Wang

Tuesday 16th October, 2018

1 Introduction

This note gives a review for Bayes factors for linear regression.

2 Mixture of g prior

This section is adapted from Liang et al. (2008). Suppose $\mathbf{Y} \in \mathbb{R}^n$ is generated from the model

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \phi^{-1}\mathbf{I}_n)$.

Let $\mathbf{X}_\gamma \in \mathbb{R}^{n\times p_\gamma}$ be a submatrix of $\mathbf{X}$. Then the submodel $\mathcal{M}_\gamma$ is defined as

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma + \boldsymbol{\varepsilon}.$$

The null model $\mathcal{M}_N$ is

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n \alpha + \boldsymbol{\varepsilon}.$$

We would like to compare $\mathcal{M}_\gamma$ with $\mathcal{M}_N$. Without loss of generality, we assume $\mathbf{1}_n^\top \mathbf{X}_\gamma = 0$. Under $\mathcal{M}_N$, the g prior is

$$p(\alpha, \phi|\mathcal{M}_N) = \frac{1}{\phi}.$$

Under $\mathcal{M}_\gamma$, the g prior is

$$\boldsymbol{\beta}_\gamma|\phi, \mathcal{M}_\gamma \sim \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1}), \quad p(\alpha|\phi, \mathcal{M}_\gamma) \propto 1, \quad p(\phi|\mathcal{M}_\gamma) = \frac{1}{\phi}.$$

The joint pdf is

$$\begin{aligned}
&p(\mathbf{Y}, \alpha, \boldsymbol{\beta}_\gamma, \phi|\mathcal{M}_\gamma) = p(\mathbf{Y}|\alpha, \boldsymbol{\beta}_\gamma, \phi, \mathcal{M}_\gamma)p(\boldsymbol{\beta}_\gamma|\phi, \mathcal{M}_\gamma)p(\alpha|\phi, \mathcal{M}_\gamma)p(\phi|\mathcal{M}_\gamma)\\
=&(2\pi)^{-(n+p_\gamma)/2} g^{-p_\gamma/2}\phi^{(n+p_\gamma)/2-1}|\mathbf{X}_\gamma^\top \mathbf{X}_\gamma|^{1/2} \exp\left\{-\frac{n\phi}{2}(\bar{\mathbf{Y}} - \alpha)^2\right\}\\
&\exp\left\{-\frac{\phi(g+1)}{2g}\left\|\mathbf{X}_\gamma\left(\boldsymbol{\beta}_\gamma - \frac{g}{g+1}\hat{\boldsymbol{\beta}}_\gamma\right)\right\|^2 - \frac{\phi}{2(g+1)}\left\|\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma\right\|^2 - \frac{\phi}{2}\left\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}} - \mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma\right\|^2\right\},
\end{aligned}$$

where $\bar{\mathbf{Y}} = n^{-1}\mathbf{1}_n^\top \mathbf{Y}$, $\hat{\boldsymbol{\beta}}_\gamma = (\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1}\mathbf{X}_\gamma^\top \mathbf{Y}$.

Direct calculation yields

$$p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \frac{\Gamma((n-1)/2)}{\pi^{(n-1)/2}\sqrt{n}} \left\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}\right\|^{-(n-1)} \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}},$$

where $R_\gamma^2 = 1 - \|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}} - \mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma\|^2 / \|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}\|^2$. Also, we have

$$p(\mathbf{Y}|\mathcal{M}_N) = \frac{\Gamma((n-1)/2)}{\pi^{(n-1)/2}\sqrt{n}} \left\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}\right\|^{-(n-1)}.$$

Thus,

$$\mathrm{BF}[\mathcal{M}_\gamma : \mathcal{M}_N] = (1+g)^{(n-p_\gamma-1)/2}[1+g(1-R_\gamma^2)]^{-(n-1)/2}.$$

2.1 Choices of g

Local empirical Bayes. The local EB estimates a separate g for each model $\mathcal{M}_\gamma$.

$$\hat{g}_\gamma^{\mathrm{EBL}} = \arg\max_{g\geq 0} p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \arg\max_{g\geq 0} \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}} = \max\{F_\gamma - 1, 0\},$$

where

$$F_\gamma = \frac{R_\gamma^2/p_\gamma}{(1-R_\gamma^2)/(n-1-p_\gamma)}$$

is the usual F statistic for testing $\boldsymbol{\beta}_\gamma = 0$.

Global empirical Bayes. The global EB procedure assumes one common g for all models.

$$\hat{g}_\gamma^{\mathrm{EBG}} = \arg\max_{g\geq 0} \sum_\gamma p(\mathcal{M}_\gamma) p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \arg\max_{g\geq 0} \sum_\gamma p(\mathcal{M}_\gamma) \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}}.$$

In general, this marginal likelihood is not tractable and does not provide a closed-form solution for $\hat{g}_\gamma^{\mathrm{EBG}}$. It can be computed by an EM algorithm, which is based on treating both the model indicator and the precision ϕ as latent data.

2.2 Mixtures of g priors

Under $\mathcal{M}_\gamma$, the mixtures of g prior take the form

$$\boldsymbol{\beta}_\gamma | g, \phi, \mathcal{M}_\gamma \sim \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1}), \quad \pi(g), \quad p(\alpha|\phi, \mathcal{M}_\gamma) \propto 1, \quad p(\phi|\mathcal{M}_\gamma) = \frac{1}{\phi}.$$

Zellner-Siow Priors

$$\pi(\boldsymbol{\beta}_\gamma|\phi) \propto \frac{\Gamma(p_\gamma)}{\pi^{p_\gamma/2}} \left|\frac{\mathbf{X}_\gamma^\top \mathbf{X}_\gamma}{n/\phi}\right|^{1/2} \left(1 + \boldsymbol{\beta}_\gamma^\top \frac{\mathbf{X}_\gamma^\top \mathbf{X}_\gamma}{n/\phi} \boldsymbol{\beta}_\gamma\right)^{-p_\gamma/2}$$

The Zellner-Siow priors can be represented as a mixture of g priors with an Inv-Gamma(1/2,n/2) prior on g, namely,

$$\phi(\boldsymbol{\beta}_\gamma|\phi) \propto \int \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1})\pi(g)dg,$$

with

$$\pi(g) = \frac{(n/2)^{1/2}}{\Gamma(1/2)} g^{-3/2} e^{-n/(2g)}.$$

Hyper-g priors

$$\pi(g) = \frac{a-2}{2}(1+g)^{-a/2}\mathbf{1}_{(0,\infty)}(g), \quad a > 2.$$

Equivalently,

$$\frac{g}{1+g} \sim \text{Beta}(1, \frac{a}{2} - 1).$$

The null-based Bayes factor is

$$\begin{aligned}\text{BF}[\mathcal{M}_\gamma : \mathcal{M}_N] =& \frac{a-2}{2}\int_0^\infty (1+g)^{(n-1-p_\gamma-a)/2}[1+(1-R_\gamma^2)g]^{-(n-1)/2}dg \\ =& \frac{a-2}{p_\gamma + a - 2} \times {}_2F_1\left(\frac{n-1}{2}, 1; \frac{p_\gamma + a}{2}; R_\gamma^2\right),\end{aligned}$$

where ${}_2F_1(a,b;c;z)$ is the Gaussian hypergeometric function defined as

$${}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a}dt.$$

Beta prime prior Maruyama and George (2011) proposed to use the beta prime prior for g:

$$\pi(g) = \frac{g^b(1+g)^{-a-b-2}}{B(a+1,b+1)}\mathbf{1}_{(0,\infty)}(g),$$

where $a > -1$, $b > -1$. Equivalently,

$$\frac{1}{1+g} \sim \text{Be}(a+1, b+1).$$

They observed that the Bayes factor has a closed form if we take

$$b = \frac{n - p_\gamma - 5}{2} - a.$$

3 Intrinsic prior

4 Normal-inverse-gamma (NIG) prior

Zhou and Guan (2018)

Consider the testing problem in linear regression with independent normal errors:

$$\begin{aligned} H_0 &: \mathbf{Y}|\mathbf{a}, \tau \sim \mathcal{N}(\mathbf{Wa}, \tau^{-1}\mathbf{I}_n), \\ H_1 &: \mathbf{Y}|\mathbf{a}, \mathbf{b}, \tau \sim \mathcal{N}(\mathbf{Wa} + \mathbf{Lb}, \tau^{-1}\mathbf{I}_n),\end{aligned}$$

where $\mathbf{W}$ is a full-rank $n \times q$ matrix representing the nuisance covariates, including a column of $\mathbf{1}_n$. $\mathbf{L}$ is an $n \times p$ matrix representing the covariates of interest.

NIG prior:

$$\begin{aligned}
\mathbf{a}|\tau &\sim \mathcal{N}(0, \tau^{-1}\mathbf{V}_a),\\
\mathbf{b}|\tau &\sim \mathcal{N}(0, \tau^{-1}\mathbf{V}_b),\\
\tau &\sim \mathrm{Gamma}(\kappa_1/2, 2/\kappa_2).
\end{aligned}$$

Here

$$\pi(\tau) = \frac{(\kappa_2/2)^{\kappa_1/2}}{\Gamma(\kappa_1/2)}\tau^{\kappa_1/2-1}\exp\left\{-\frac{\kappa_2\tau}{2}\right\}$$

Then

$$\begin{aligned}
&f(\mathbf{Y}|\mathbf{a},\mathbf{b},\tau)\pi(\mathbf{a}|\tau)\pi(\mathbf{b}|\tau)\pi(\tau)\\
=&\frac{(\kappa_2/2)^{\kappa_1/2}\tau^{(n+p+q+\kappa_1)/2-1}}{(2\pi)^{(n+p+q)/2}|\mathbf{V}_a|^{1/2}|\mathbf{V}_b|^{1/2}\Gamma(\kappa_1/2)}\exp\left\{-\frac{\tau}{2}\left(\|\mathbf{Y}-\mathbf{W}\mathbf{a}-\mathbf{L}\mathbf{b}\|^2+\mathbf{a}^\top\mathbf{V}_a^{-1}\mathbf{a}+\mathbf{b}^\top\mathbf{V}_b^{-1}\mathbf{b}+\kappa_2\right)\right\}.
\end{aligned}$$

Let $\mathbf{H}_\mathbf{W} = \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top$. Then

$$\begin{aligned}
\|\mathbf{Y}-\mathbf{W}\mathbf{a}-\mathbf{L}\mathbf{b}\|^2 =&\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2+\|\hat{\mathbf{Y}}-\mathbf{W}\mathbf{a}-\mathbf{L}\mathbf{b}\|^2\\
=&\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2+\|\mathbf{H}_\mathbf{W}(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})-\mathbf{W}\mathbf{a}\|^2+\|(\mathbf{I}_n-\mathbf{H}_\mathbf{W})(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})\|^2.
\end{aligned}$$

We have

$$\begin{aligned}
&\int_{\mathbb{R}^q}\exp\left\{-\frac{\tau}{2}\left(\|\mathbf{H}_\mathbf{W}(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})-\mathbf{W}\mathbf{a}\|^2+\mathbf{a}^\top\mathbf{V}_a^{-1}\mathbf{a}\right)\right\}d\mathbf{a}\\
=&\int_{\mathbb{R}^q}\exp\Big\{-\frac{\tau}{2}\Big(\Big\|(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{1/2}\Big[\mathbf{a}-(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})\Big]\Big\|^2\\
&-(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})^\top\mathbf{W}(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})+\|\mathbf{H}_\mathbf{W}(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})\|^2\Big)\Big\}d\mathbf{a}\\
=&(2\pi)^{q/2}\tau^{-q/2}|\mathbf{W}\mathbf{W}^\top+\mathbf{V}_a^{-1}|^{-1/2}\\
&\exp\left\{-\frac{\tau}{2}\left(-(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})^\top\mathbf{W}(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})+\|\mathbf{H}_\mathbf{W}(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})\|^2\right)\right\}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\int_{\mathbb{R}^q}f(\mathbf{Y}|\mathbf{a},\mathbf{b},\tau)\pi(\mathbf{a}|\tau)\pi(\mathbf{b}|\tau)\pi(\tau)d\mathbf{a}\\
=&\frac{(\kappa_2/2)^{\kappa_1/2}\tau^{(n+p+\kappa_1)/2-1}}{(2\pi)^{(n+p)/2}|\mathbf{V}_a|^{1/2}|\mathbf{V}_b|^{1/2}\Gamma(\kappa_1/2)|\mathbf{W}\mathbf{W}^\top+\mathbf{V}_a^{-1}|^{1/2}}\\
&\exp\Big\{-\frac{\tau}{2}\Big(\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2-(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})^\top\mathbf{W}(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})+\|\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b}\|^2\\
&+\mathbf{b}^\top\mathbf{V}_b^{-1}\mathbf{b}+\kappa_2\Big)\Big\}.
\end{aligned}$$

Note that

$$\begin{aligned}
&\int_{\mathbb{R}^p}\exp\left\{-\frac{\tau}{2}\left(-(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})^\top\mathbf{W}(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})+\|\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b}\|^2+\mathbf{b}^\top\mathbf{V}_b^{-1}\mathbf{b}\right)\right\}d\mathbf{b}\\
&=(2\pi)^{p/2}\tau^{-p/2}|\mathbf{L}^\top\mathbf{C}\mathbf{L}+\mathbf{V}_b^{-1}|^{-1/2}\exp\left\{-\frac{\tau}{2}\hat{\mathbf{Y}}^\top\left(\mathbf{C}-\mathbf{C}\mathbf{L}(\mathbf{L}^\top\mathbf{C}\mathbf{L}+\mathbf{V}_b^{-1})^{-1}\mathbf{L}^\top\mathbf{C}\right)\hat{\mathbf{Y}}\right\},
\end{aligned}$$

where $\mathbf{C} = \mathbf{I}_n - \mathbf{W}(\mathbf{W}^\top\mathbf{W} + \mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top$. Thus,

$$
\begin{aligned}
&\int_{\mathbb{R}^p}\int_{\mathbb{R}^q} f(\mathbf{Y}|\mathbf{a},\mathbf{b},\tau)\pi(\mathbf{a}|\tau)\pi(\mathbf{b}|\tau)\pi(\tau)d\mathbf{a}d\mathbf{b}\\
=&\frac{(\kappa_2/2)^{\kappa_1/2}\tau^{(n+\kappa_1)/2-1}}{(2\pi)^{n/2}|\mathbf{V}_a|^{1/2}|\mathbf{V}_b|^{1/2}\Gamma(\kappa_1/2)|\mathbf{W}\mathbf{W}^\top + \mathbf{V}_a^{-1}|^{1/2}|\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1}|^{1/2}}\\
&\exp\Big\{-\frac{\tau}{2}\Big(\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2 + \hat{\mathbf{Y}}^\top\left(\mathbf{C} - \mathbf{C}\mathbf{L}(\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1})^{-1}\mathbf{L}^\top\mathbf{C}\right)\hat{\mathbf{Y}} + \kappa_2\Big)\Big\}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\int_0^\infty\int_{\mathbb{R}^p}\int_{\mathbb{R}^q} f(\mathbf{Y}|\mathbf{a},\mathbf{b},\tau)\pi(\mathbf{a}|\tau)\pi(\mathbf{b}|\tau)\pi(\tau)d\mathbf{a}d\mathbf{b}d\tau\\
=&\frac{(\kappa_2/2)^{\kappa_1/2}\Gamma((n+\kappa_1)/2)}{(2\pi)^{n/2}|\mathbf{V}_a|^{1/2}|\mathbf{V}_b|^{1/2}\Gamma(\kappa_1/2)|\mathbf{W}\mathbf{W}^\top + \mathbf{V}_a^{-1}|^{1/2}|\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1}|^{1/2}}\\
&\left(\frac{\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2 + \hat{\mathbf{Y}}^\top\left(\mathbf{C} - \mathbf{C}\mathbf{L}(\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1})^{-1}\mathbf{L}^\top\mathbf{C}\right)\hat{\mathbf{Y}} + \kappa_2}{2}\right)^{-(n+\kappa_1)/2}.
\end{aligned}
$$

Under H_0, that is $\mathbf{V}_b \to 0$, we have

$$
\begin{aligned}
&\int_0^\infty\int_{\mathbb{R}^q} f(\mathbf{Y}|\mathbf{a},0,\tau)\pi(\mathbf{a}|\tau)\pi(\tau)d\mathbf{a}d\tau\\
=&\frac{(\kappa_2/2)^{\kappa_1/2}\Gamma((n+\kappa_1)/2)}{(2\pi)^{n/2}|\mathbf{V}_a|^{1/2}\Gamma(\kappa_1/2)|\mathbf{W}\mathbf{W}^\top + \mathbf{V}_a^{-1}|^{1/2}}\\
&\left(\frac{\|\mathbf{Y}-\hat{\mathbf{Y}}_0\|^2 + \hat{\mathbf{Y}}_0^\top\mathbf{C}\hat{\mathbf{Y}}_0 + \kappa_2}{2}\right)^{-(n+\kappa_1)/2}.
\end{aligned}
$$

Thus,

$$
\mathrm{BF} = \frac{1}{|\mathbf{V}_b|^{1/2}|\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1}|^{1/2}}\left(\frac{\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2 + \hat{\mathbf{Y}}^\top\left(\mathbf{C} - \mathbf{C}\mathbf{L}(\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1})^{-1}\mathbf{L}^\top\mathbf{C}\right)\hat{\mathbf{Y}} + \kappa_2}{\|\mathbf{Y}-\hat{\mathbf{Y}}_0\|^2 + \hat{\mathbf{Y}}_0^\top\mathbf{C}\hat{\mathbf{Y}}_0 + \kappa_2}\right)^{-(n+\kappa_1)/2}
$$

References

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Maruyama, Y. and George, E. I. (2011). Fully bayes factors with a generalized g -prior. *Ann. Statist.*, 39(5):2740–2765.

Zhou, Q. and Guan, Y. (2018). On the null distribution of bayes factors in linear regression. *Journal of the American Statistical Association*, 0(0):1–10.